

Rhode Island utilizes novel approaches to analytics including automated data processing



CONTRIBUTOR: Sarah Bowman, COVID-19 Chief Data Officer, Rhode Island Department of Health

CATEGORY: **Epidemiology and Laboratory Capacity (ELC)**

CATEGORY: **Interoperable Data Systems**

The Rhode Island Department of Health implemented a data lake to aggregate and process data from multiple sources. This streamlined COVID-19 data analysis efforts and supported dashboard creation. Moving forward, this resource may be leveraged to integrate new disease-specific components.

The “What”

The COVID-19 surveillance team was responsible for managing large and rapidly expanding data sets to produce an unprecedented volume of analyses and routine metrics. Namely, advanced data management processes were required to synthesize the many incoming lab results, identify which were for the same individual, and to identify new cases among the many positive results. This effort was further complicated by testing strategies that



included concurrent rapid and PCR testing and cyclic testing within congregate care settings as well as testing scheduled directly by community members, both of which resulted in multiple results for the same person, and often multiple positive results associated with a single infection. Demographics may have varied in content or completeness across records adding to the complexity of data management. This foundational data processing was critical to support reporting on routine case counts, and the underlying testing related metrics that were essential for routine monitoring throughout the response.



More stories are available at stories.cste.org

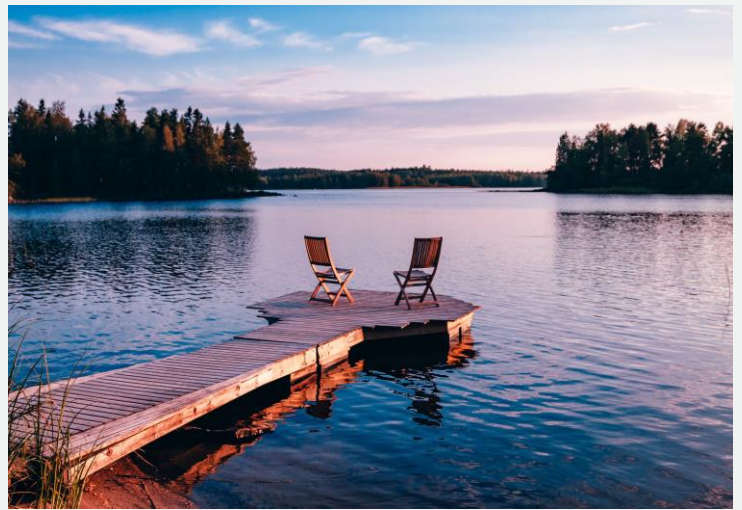
SUBMITTED JUNE 2023

The Intermediary Data Layer (IDL) data lake is a solution that aggregates and processes data from the back end of Salesforce and other sources, providing analysts with a centralized location of data for a growing number of COVID-19 surveillance areas.

Hospitalization, and death data sets. The IDL provides a data management solution that consolidates data from multiple sources to offer a single access point for data sets spanning the COVID-19 surveillance effort and recently expanded to incorporate wastewater and emergency department (ED) data as well. The use case for the IDL has expanded to support routine output that is leveraged to support the production of weekly and monthly web and dashboard content. Development and maintenance of the routine output produced by the IDL moved routine data processes out of the daily workflow for the epidemiologists, allowing this team to focus on more nuanced and ad hoc analyses.

The “So What”

The Intermediary Data Layer (IDL) data lake is a solution that aggregates and processes data from the back end of Salesforce and other sources, providing analysts with a centralized location of data for a growing number of COVID-19 surveillance areas. A central function of the IDL is the underlying data processing to maintain the foundational case data set, and to establish demographics that pull from many sources (within and beyond Salesforce) to report the most complete and consistent demographics across records for people appearing in the case,




The “Now What”

Rhode Island Department of Health (RIDOH) continues to use the IDL to manage and maintain the foundational COVID-19 data sets and has been able to expand on the initial use case of the IDL to support the data processing for routine output as well. RIDOH is now undertaking an assessment of the IDL, as it was utilized as part of the COVID-19 response, to determine opportunities for expanding

its scope to include additional diseases. By leveraging the existing framework, RIDOH aims to apply lessons learned and extend the benefits by seamlessly integrating new disease-specific components without disrupting the functionality of the current IDL system to support ongoing COVID-19 analytics.

Key contributors to this project include the COVID-19 Surveillance Team (Quant team), Rhode Island Department of Health, and IBM Analytics Team.

 Funding source: This work was supported in part by the Centers for Disease Control and Prevention (CDC) cooperative agreement #NU38OT000297. Its contents are solely the responsibility of the authors and do not necessarily reflect the views of the CDC.